

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平6-332782

(43)公開日 平成6年(1994)12月2日

(51)Int.Cl. <sup>5</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 12/00	5 4 5 B	8944-5B		
15/16	3 7 0 M	7429-5L		
	3 8 0 Z	7429-5L		

審査請求 未請求 請求項の数21 O L (全 21 頁)

(21)出願番号 特願平6-50126  
 (22)出願日 平成6年(1994)3月22日  
 (31)優先権主張番号 特願平5-61602  
 (32)優先日 平5(1993)3月22日  
 (33)優先権主張国 日本 (J P)

(71)出願人 000005108  
 株式会社日立製作所  
 東京都千代田区神田駿河台四丁目6番地  
 (71)出願人 000233011  
 日立コンピュータエンジニアリング株式会  
 社  
 神奈川県秦野市堀山下1番地  
 (72)発明者 秋沢 充  
 東京都国分寺市東恋ヶ窪1丁目280番地  
 株式会社日立製作所中央研究所内  
 (74)代理人 弁理士 小川 勝男

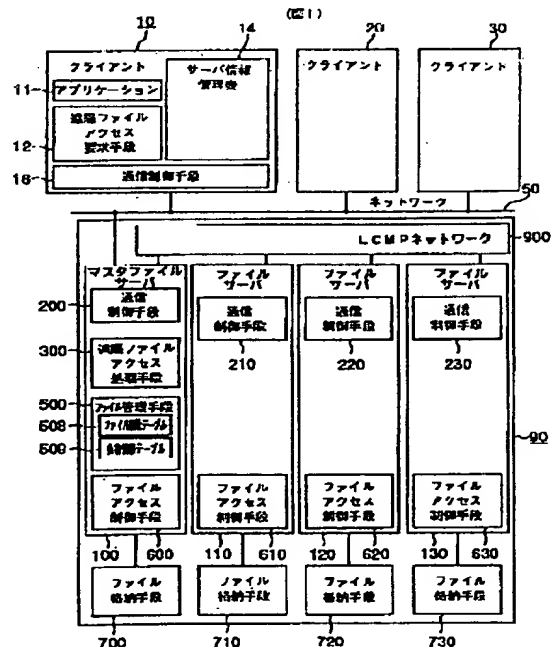
最終頁に続く

(54)【発明の名称】 ファイルサーバシステム及びそのファイルアクセス制御方法

(57)【要約】

【目的】 各々のファイル格納装置をアクセスする複数のファイルサーバがネットワークを介して並設されたファイルサーバシステムにおいて、複数のクライアントから特定のファイルサーバにアクセス要求が集中することによるスレープットの低下を防ぐ。

【構成】 ファイルサーバシステム90を構成するファイルサーバ100、110、120、130のうちのマスタファイルサーバ100には、各ファイルサーバの負荷状況を計測・管理する負荷情報テーブル508とファイルブロックごとのアクセス担当のファイルサーバを記録・管理するファイル属性テーブルとを用いてファイルを管理し、とくにファイル書き込み時には負荷の軽いファイルサーバを選定し、選定されたファイルサーバにクライアント計算機10、20、30から伝送されたファイルアクセス要求を分配するファイル管理手段を備える。



【特許請求の範囲】

【請求項1】 ネットワーク上に並接された複数のファイルサーバを有し、複数のクライアント計算機間で上記ファイルサーバに分散配置されたファイルを共有するファイルサーバシステムにおいて、

上記複数のファイルサーバの各々に、

ファイルを格納するファイル格納手段と、

上記ネットワークを介して他のファイルサーバとの通信制御を行う第一の通信制御手段と、

ファイルアクセス要求を受け付けて上記ファイル格納手段に対してファイルアクセスを行なうファイルアクセス制御手段を設け、

上記複数のファイルサーバのうち特定のファイルサーバに、

上記クライアント計算機との通信制御を行なう第二の通信制御手段と、

上記クライアント計算機から発行されたファイルアクセス要求の通信プロトコルを管理する遠隔ファイルアクセス処理手段と、

上記複数のファイルサーバの各々の負荷状況を計測する負荷情報モニタリング手段と、

上記負荷情報モニタリング手段によって計測した負荷状況を参照して上記複数のファイルサーバからファイルアクセスを行なうファイルサーバを選定し、選定されたファイルサーバが自己のファイルサーバであるときに自己のファイルサーバのファイルアクセス制御手段に対してファイルアクセス要求を発行し、選定されたファイルサーバが他のファイルサーバであるときに上記第1の通信制御手段を介してその選定されたファイルサーバのファイルアクセス制御手段に対してファイルアクセス要求を発行するファイルアクセス要求を発行するファイルアクセス要求配分手段を更に設けたことを特徴とするファイルサーバシステム。

【請求項2】 上記負荷情報モニタリング手段は、上記複数のファイルサーバの各々における未処理のファイルアクセス要求数を計数する手段を含む請求項1に記載のファイルサーバシステム。

【請求項3】 上記ファイルアクセス要求配分手段は、上記クライアント計算機から発行されたファイルアクセス要求が書き込み要求か読み出し要求かを判定する書き込み読み出し判定手段と、

ファイル書き込み時には前記負荷情報モニタリング手段によって計測した負荷状況を参照してファイルを格納するファイルサーバを選定するファイル分割配置手段と、ファイル読み出し時には上記負荷情報モニタリング手段によって計測した負荷状況を参照して読み出し対象ファイルが格納されているファイルサーバから読み出しの対象とするファイルサーバを選定するアクセス対象ファイルサーバ・スケジューリング手段を含むことを特徴とする請求項1に記載のファイルサーバシステム。

【請求項4】 上記ファイル分割配置手段は、書き込み対象のファイルを格納するファイルサーバを少なくとも二つ以上選定するファイル分割配置手段を設けたことを特徴とする請求項3に記載のファイルサーバシステム。

【請求項5】 上記ファイル分割配置手段は、書き込み対象のファイルを格納するファイルサーバを少なくとも二つ以上選定し、上記アクセス対象ファイルサーバ・スケジューリング手段は読み出し対象のファイルが格納されているファイルサーバのうち前記負荷情報モニタリング手段により取得された負荷状況を参照し負荷の低いファイルサーバを読み出し対象のファイルサーバとして選定することを特徴とする請求項3に記載のファイルサーバシステム。

【請求項6】 上記ファイル分割配置手段は、ファイルを格納するファイルサーバを選定する際に、該ファイルと該ファイルサーバの対応関係を示すファイル属性テーブルを生成し、上記アクセス対象サーバスケジューリング手段は上記ファイル属性テーブルを参照して読み出し対象のファイルが格納されているファイルサーバを選定することを特徴とする請求項3に記載のファイルサーバシステム。

【請求項7】 上記第二の通信制御手段と上記遠隔ファイルアクセス処理手段とを上記複数のファイルサーバのうちの少なくとも二つ以上にそれぞれ設けたことを特徴とする請求項1に記載のファイルサーバシステム。

【請求項8】 上記負荷情報モニタリング手段と上記ファイルアクセス要求配分手段とを上記複数のファイルサーバのうちの少なくとも二つ以上にそれぞれ設けたことを特徴とする請求項1に記載のファイルサーバシステム。

【請求項9】 上記負荷情報モニタリング手段と上記ファイルアクセス要求配分手段とを上記複数のファイルサーバのうちの少なくとも二つ以上にそれぞれ設けたことを特徴とする請求項7に記載のファイルサーバシステム。

【請求項10】 上記クライアント計算機は上記ネットワークに接続され、上記クライアント計算機と上記特定のファイルサーバとの通信制御の機能は上記第二の通信制御手段の代りに上記第一の通信制御手段が果たすことを特徴とする請求項1に記載のファイルサーバシステム。

【請求項11】 上記クライアント計算機は第1のネットワークに接続され、上記複数のファイルサーバは第2のネットワークに接続され、上記第1、第2のネットワークは上記クライアント計算機からのファイルアクセス要求を上記特定のファイルサーバへ配分するブリッジ手段で接続されることを特徴とする請求項1に記載のファイルサーバシステム。

【請求項12】 上記第一の通信制御手段はシステムバスを介して他のファイルサーバとの通信を行うことを特徴とする請求項1に記載のファイルサーバシステム。

【請求項13】 上記システムバスは上記複数のファイルサーバ間の通信のみに用いる専用バスであることを特徴

とする請求項12に記載のファイルサーバシステム。

【請求項14】ネットワーク上に並接された複数のファイルサーバを有し、複数のクライアント計算機間で上記ファイルの複数のファイルサーバに分散配置されたファイルを共有するファイルサーバシステムにおけるファイルアクセス制御方法において、

上記複数のファイルサーバの各々の負荷情報を計測し、クライアント計算機から上記ネットワークを介して発行されたファイルアクセス要求が受け付けられた際に、上記負荷情報を参照してファイルアクセスを行うファイルサーバを選定し、

上記選定ファイルサーバに対してファイルアクセス要求を配分する、

とのステップを含むファイルアクセス制御方法。

【請求項15】上記の負荷情報を計測するステップは上記複数のファイルサーバの各々の未処理のファイルアクセス要求数を計数するステップを含む請求項14に記載のファイルアクセス制御方法。

【請求項16】上記ファイルアクセスを行なうファイルサーバを選定するステップは、クライアント計算機から発行されたファイルアクセス要求が書き込み要求か読み出し要求かを判定し、ファイル書き込み時には計測した負荷状況を参照してファイル格納対象のファイルサーバを選定し、ファイル読み出し時には計測した負荷状況を参照して読み出し対象ファイルが格納されているファイルサーバから読み出し対象のファイルサーバを選定する、とのステップを含むことを特徴とする請求項14に記載のファイルアクセス制御方法。

【請求項17】請求項16に記載のファイルアクセス制御方法において、上記のファイル格納対象のファイルサーバを選定するステップでは書き込み対象ファイルを格納するファイルサーバを少なくとも二つ以上選定することを特徴とするファイルアクセス制御方法。

【請求項18】請求項16に記載のファイルアクセス制御方法において、上記の読み出し対象のファイルサーバを選定するステップでは読み出し対象ファイルが格納されている複数のファイルサーバのうち計測した負荷情報に基づき負荷の軽いファイルサーバを選択して読み出し対象のファイルサーバとすることを特徴とするファイルアクセス制御方法。

【請求項19】請求項16に記載のファイルアクセス制御方法において、ファイルを格納するファイルサーバを選定する際に、該ファイルと該ファイルサーバの対応関係を示すファイル属性テーブルを生成し、上記アクセス対象サーバスケジューリング手段は上記ファイル属性テーブルを参照して読み出し対象のファイルが格納されているファイルサーバを特定する

【請求項20】請求項14に記載のファイルアクセス制御方法において、上記ファイルアクセス対象ファイルサ

ーバを選定するステップは、

少なくとも二つ以上のファイルサーバで各ファイルサーバの負荷情報を計測し、

クライアント計算機から発行されたファイルアクセス要求を上記負荷情報を計測するファイルサーバで受け取り、

上記負荷情報を参照してファイルアクセス対象ファイルサーバを選定し、

ファイルアクセス対象ファイルサーバとして選定したファイルサーバに対してファイルアクセス要求を配分することを特徴とする高速ファイルアクセス制御方法。

【請求項21】請求項20に記載のファイルアクセス制御方法において、上記ファイルアクセス対象ファイルサーバを選定するステップでは、ファイルアクセスがファイル書き込みである時には負荷の少ないファイルサーバを書き込み対象ファイルサーバとして選定し、ファイルアクセスがファイル読み出しである時にはファイルが格納されているファイルサーバから負荷の少ないファイルサーバを読み出し対象ファイルサーバとして選定することを特徴とするファイルアクセス制御方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明はワークステーションやサーバ等の計算機システムに関わり、特にマルチプロセッサ構成の計算機システムにおける、二次記憶装置に格納されたファイルを高速にアクセスするファイルサーバシステム及びそのファイルアクセス制御方法に関する。

【0002】

【従来の技術】近年、計算機のネットワーク化が進展してきている。これに伴い、計算機間で共有するファイルを一括して管理するファイルサーバの需要が高まっている。これは、低コストでファイルシステムを構築できるためである。すなわち、ファイルサーバを用いることによって複数の計算機間でファイルの共有が可能となるため、同一のファイルを複数の計算機間で複数コピーして所持しなくても済むようになるからである。ファイルサーバには通常ネットワーク対応ファイルシステムが搭載されており、同じネットワークに接続されたクライアント側の計算機にもネットワーク対応ファイルシステム・アクセスプログラムを搭載することによって、あたかもクライアント計算機自身に格納されているファイルであるかのようにアクセスすることが可能になる。そのため、ネットワークに接続されているどのクライアント計算機からも、ファイルサーバ上に蓄積、管理されているファイルに対してアクセスすることが可能となり、複数のクライアント計算機間でのファイルの共有が実現される。

【0003】ネットワーク対応ファイルシステムについての記述は、Managing NFS and NIS (Hal Stern, O'Reilly & Associates, Inc., June 1991, p. 113 ~ p. 159)にある。

【0004】しかし、ファイルサーバ内のファイルをネットワーク対応ファイルシステムを用いて共有する場合には、性能上の問題が生じることがある。すなわち多数のクライアント計算機から同時にファイルアクセスを行なう場合には、ファイルサーバに負荷が集中し、直ちにアクセス結果が得られない状況が発生する。

【0005】そのため、ファイルサーバへの負荷が過大になりクライアント計算機のアクセス・スループットが低下する場合には、複数台のファイルサーバをネットワーク上に並接し複数クライアント計算機からのアクセス要求を分散することが行なわれている。各クライアント計算機は、ネットワーク上に設置されている全ファイルサーバの情報を管理し、この管理情報を参照してアクセス対象のファイルが存在するファイルサーバに対してアクセス要求を発行する。これにより、各クライアント計算機のアクセス対象ファイルが別々のファイルサーバに格納されている場合には、負荷を分散することが可能になり、アクセススループットを向上することができる。

【0006】クライアント・アンド・サーバ構成ではないが、一つのメインプロセッサが複数のデータファイルの全体の管理を行い、これらデータ・ファイルは複数のサブ・プロセッサに分散配置されているとの点で上記システムと類似するシステムが日本公開特許04-48352号(1992)に記載される。

【0007】

【発明が解決しようとする課題】上述したように複数台のファイルサーバを設置して多数のクライアント計算機間でファイルを共有したとしても、以下の問題が残る。

【0008】クライアント計算機はファイルサーバの負荷状況とは全く無関係にファイルが存在するファイルサーバをアクセスするため、複数のクライアント計算機が一つのファイルサーバに同時にアクセス要求を出すことがあり、その場合には該当ファイルサーバがボトルネックとなり、スループットが低下してしまうという問題が生じる。とくに、複数のクライアント計算機が同一のディレクトリや同一のファイルに同時にアクセスすると、その性能低下が甚だしくなる。この問題はその性格上ファイルサーバの設置台数を増やしても解決できるものではない。

【0009】本発明の目的は、多数のクライアント計算機間でのファイルの共有を行なうために複数のファイルサーバを並接したネットワークにおいて、複数のクライアント計算機が同一のディレクトリやファイルに同時にアクセスした場合でも、特定のファイルサーバへのアクセスの集中によるボトルネックの発生とそれに伴うスループットの低下を防ぐことができるファイルサーバシステムとそのファイルアクセス制御方法を提供することである。

【0010】

【課題を解決するための手段】本発明のファイルアクセ

ス制御方法の一つの特徴は、ネットワーク上に並接された複数のファイルサーバを有し、複数のクライアント計算機間で複数のファイルサーバに分散配置されたファイルを共有するファイルサーバシステムにおいて、上記複数のファイルサーバの各々の負荷情報を計測し、クライアント計算機から上記ネットワークを介して発行されたファイルアクセス要求が受け付けられた際に、上記負荷情報を参照してファイルアクセスを行うファイルサーバを選定し、上記選定ファイルサーバに対してファイルアクセス要求を配分する、とのステップを含むファイルアクセス制御方法にある。

【0011】より具体的にいえば、クライアント計算機からのファイルアクセス要求が新たなファイルの書き込み要求である場合には複数のファイルサーバの負荷情報を計測して負荷の最も軽いファイルサーバを選定し、そのファイルサーバにファイル書き込み要求を発行する。あるいは、ミラー構成のファイルを作製する場合には、負荷の軽い複数のファイルサーバを選定し、それらのファイルサーバそれぞれにファイル書き込み要求を発行する。また、クライアント計算機からのファイルアクセス要求がファイル読み出し要求であり、読み出し対象のファイルがミラー構成にされている場合には、読み出し対象のファイルが格納された複数の複数のファイル格納装置それぞれを受け持つ複数のファイルサーバの負荷情報を計測して負荷の最も軽いひとつのファイルサーバを選定し、そのファイルサーバにファイル読み出し要求を発行する。

【0012】上記の負荷情報の計測は、上記複数のファイルサーバの各々の未処理のファイルアクセス要求数を計数することにより行う。このために、複数のファイルサーバのうちの少なくとも一つをマスタ・ファイルサーバとし、このマスタ・ファイルサーバには、複数のファイルサーバの各々の未処理のファイルアクセス要求数、負荷情報テーブルに記録・更新する負荷情報モニタリング手段を設ける。また、このマスタ・ファイルサーバには、各ファイルとそのファイルの書き込みを受け持ったファイルサーバの対応関係を記録するファイル属性テーブル、および上記負荷情報テーブルを用いてクライアント計算機からのファイルアクセス要求を配分する手段を設ける。

【0013】本発明にしたがうファイルサーバシステムの代表的構成は、ネットワーク上に並接された複数のファイルサーバを有し、複数のクライアント計算機間で上記ファイルサーバに分散配置されたファイルを共有するファイルサーバシステムにおいて、上記複数のファイルサーバの各々には、ファイルを格納するファイル格納手段と、上記ネットワークを介して他のファイルサーバとの通信制御を行う第一の通信制御手段と、ファイルアクセス要求を受け付けて上記ファイル格納手段に対してファイルアクセスを行なうファイルアクセス制御手段を設

け、一方、上記複数のファイルサーバのうち特定のファイルサーバにはこれらに加えて、上記クライアント計算機との通信制御を行なう第二の通信制御手段と、上記クライアント計算機から発行されたファイルアクセス要求の通信プロトコルを管理する遠隔ファイルアクセス処理手段と、上記複数のファイルサーバの各々の負荷状況を計測する負荷情報モニタリング手段と、上記負荷情報モニタリング手段によって計測した負荷状況を参照して上記複数のファイルサーバからファイルアクセスを行なうファイルサーバを選定し、選定されたファイルサーバが自己のファイルサーバであるときに自己のファイルサーバのファイルアクセス制御手段に対してファイルアクセス要求を発行し、選定されたファイルサーバが他のファイルサーバであるときに上記第1の通信制御手段を介してその選定されたファイルサーバのファイルアクセス制御手段に対してファイルアクセス要求を発行するファイルアクセス要求配分手段を更に設けた、との構成である。

#### 【0014】

【作用】このような方法及びシステム構成によれば、ファイルアクセス負荷の少ないファイルサーバへアクセスを行なうことができる。しかも、ファイルとその複製ファイルを複数のファイルサーバに格納するため同一のディレクトリやファイルに対するクライアント計算機からのアクセス要求も複数のファイルサーバにその負荷状況に応じて分散することができる。すなわち、ネットワーク上に複数のファイルサーバを並接し多数のクライアント計算機間でファイルの共有を行なう際に、複数のクライアント計算機が同一のディレクトリやファイルに同時にアクセスした場合でも、特定のファイルサーバへのアクセスの集中によるボトルネックの発生とそれに伴うスループットの低下を防ぐことができ、クライアント計算機からの高スループットのアクセスを実現できることになる。

#### 【0015】

【実施例】本発明の実施例の構成を図1を参照して説明する。

【0016】疎結合マルチプロセッサで構成するファイルサーバシステム90と、クライアント計算器10、20、30とはローカルエリアネットワーク50により接続される。クライアント計算機10、20、30の各々ではアプリケーションプログラム11が実行され、これによってファイルアクセス要求が発生すると、遠隔ファイルアクセス要求発生手段12からファイルサーバシステム90へファイルアクセス要求が発行される。具体的には、ファイルアクセス要求は、通信制御手段16から、ファイルサーバシステム90を構成する4台のファイルサーバ100、110、120、130のうちの特定の一つであるファイルサーバ100へローカルエリアネットワーク50を介して伝送される。4台のファイル

サーバ100、110、120および130はそれぞれがファイル格納手段700、710、720および730のアクセスを個別に受け持つ。そのためにファイルサーバ100、110、120および130にはそれぞれファイルアクセス制御手段600、610、620および630が形成される。またファイルサーバ100、110、120および130はLCMPネットワーク900を介して互いに通信する。このために通信制御手段200、210、220、230がそれぞれのファイルサーバに形成される。

【0017】本実施例では、上記特定のファイルサーバ100をマスタファイルサーバと呼ぶ。マスタファイルサーバ100は、さらにクライアント計算機から発行されたファイルアクセス要求を受け付けるための遠隔ファイルアクセス処理手段300、および個々のファイルサーバの負荷が大きく偏らないようファイルの分散配置を管理し、且つ受け付けたファイルアクセス要求を個々のファイルサーバに振り分けるファイル管理手段500が形成される。このファイル管理およびファイルアクセス要求の振り分けのために、ファイル属性テーブル508および負荷情報テーブル509が用いられる。クライアント計算機のサーバ情報管理表14には、マスタファイルサーバ100の情報、つまりマスタファイルサーバ100のマシンアドレスが格納される。

【0018】図2はマスタファイルサーバ100の装置構成を示すブロック図である。マスタファイルサーバ100はシステムバス107で互いに接続されたプロセッサ101、メモリ102、ネットワークインタフェース回路103及びLCMPネットワークインタフェース回路104を含む。システム立ちあげ時に、図示しない2次記憶装置から遠隔ファイルアクセス処理プログラム301、ファイル管理プログラム501、ファイルアクセス制御プログラム601及び通信制御プログラム201がそれぞれメモリ102にロードされ、これにより図1に示した遠隔ファイルアクセス処理手段300、ファイル管理手段500、ファイルアクセス制御手段600及び通信制御手段200が形成される。マスタファイルサーバ100が書き込み読み出しを受け持つファイル格納手段700は、図2に示す通り磁気ディスク装置であり、システムバス107に接続される。なおファイル格納手段700は光磁気ディスク装置や光ディスク装置、またその他の2次記憶装置であっても構わない。

【0019】図1のマスタファイルサーバ100以外のファイルサーバ110、120及び130も、図2に示すのとほぼ同様な構成をそれぞれ有する。但し、これらの主メモリには遠隔ファイルアクセスプログラム並びにファイル管理プログラムはロードされない。またローカルエリアネットワーク50への接続のためのネットワークインタフェース回路103も不要である。

【0020】図3はマスタファイルサーバ100のプロ

グラム構成を示す。通信制御プログラム201は、ローカルエリアネットワーク50とマスタファイルサーバ100とのインタフェースとなるネットワークアクセスプログラム205と、LCMPネットワーク900とマスタファイルサーバ100とのインタフェースとなるプロセッサ間通信装置アクセスプログラム206と、ネットワークアクセスプログラム205から受け取った要求を遠隔ファイルアクセス処理プログラム301が解釈できるようにプロトコル変換して渡すネットワーク通信プロトコル制御プログラム207と、後で説明するファイルアクセス制御プログラム601で解釈された他のファイルサーバへのアクセス要求をプロトコル変換してプロセッサ間通信装置アクセスプログラム206に渡すプロセッサ間通信プロトコル制御プログラム208から構成される。ファイルアクセス制御プログラム601は、ファイル管理プログラム501からファイル格納装置およびファイルサーバに関する情報を受け取り、自己のプロセッサ101がアクセス制御を受け持つファイル格納装置700にアクセスする場合にはファイル格納装置アクセスプログラム604へファイル格納装置に関する情報を渡し、他のファイルサーバ110~130がアクセス制御を受け持つファイル格納装置710~730にアクセスする場合にはプロセッサ間通信プロトコル制御プログラム208にファイル格納装置に関する情報を渡して他のファイルサーバにアクセスを依頼するファイル格納装置識別プログラム603と、ファイル格納装置識別プログラム603からファイル格納装置に関する情報を受け取り、目的のファイルを格納する磁気ディスク装置700にアクセスするファイル格納装置アクセスプログラム604とから構成される。

【0021】ファイル管理プログラム501については、さらに詳細なプログラム構成を示す図4をも参照して説明する。ファイル管理プログラム501は、ファイル属性テーブル508と負荷情報テーブル509を管理し、またこれらを用いて遠隔ファイルアクセス処理プログラム301から渡されたファイルアクセス要求をファイルサーバの情報、ファイル格納装置の情報およびファイル格納装置内のファイル格納位置の情報に変換してファイルアクセス制御プログラム601に渡す。ファイルの分散配置及びそれにしたがうアクセス対象のファイルサーバ選定の処理をおこなう。そのために、ファイル管理プログラム501は、遠隔ファイルアクセス処理プログラム301から渡されるファイルアクセス要求を受け付けて、それが書き込み要求か読み出し要求を判別するファイルアクセス要求受付プログラム504と、ファイル書き込み時にどのファイルサーバによりファイルを書き込むかを決定するファイル分散配置プログラム502と、読み出し時にどのファイルサーバにアクセスを行なうかを決定する読み出し要求スケジューリングプログラム503と、各ファイルサーバの未処理アクセス要求数

を計数することにより各ファイルサーバの負荷状況を計測する負荷情報モニタリングプログラム505とから構成される。ファイル属性テーブル508には、各ファイルに対応しそのファイルが格納されているファイルサーバ識別子とファイル格納装置識別子およびファイル格納装置内のファイル格納位置の情報が保持される。負荷情報テーブル509には各ファイルサーバの未処理アクセス要求数が保持される。

【0022】次に、ファイル属性テーブルの第1の例を図5に示す。この例は、ひとつのファイルを分割せずに、あるいは複製を持たずに、ファイル単位での分散配置を行う場合の例である。ファイル属性テーブルは、

(1) ファイル属性領域、(2) ディスクブロックインデックス領域の二つの領域から構成される。ファイル属性領域は、ファイルサイズ、ファイル格納モード、ファイルアクセス・プロセッサ識別子、ファイル格納デバイス識別子の各エントリからなる。ファイル格納モードにはローカルとリモートがあり、ファイル属性テーブルを管理している自己のファイルサーバがアクセス制御を受けもつファイル格納装置にファイルが格納されているのか、あるいは他のファイルサーバがアクセス制御を受けもつファイル格納装置にファイルが格納されているのかを示す。ファイルアクセス・プロセッサ識別子は、ファイル属性テーブルに対応するファイルが格納されているファイル格納装置のアクセス制御を受けもつファイルサーバの識別子を示す。ファイル格納デバイス識別子は、ファイルが格納されているファイル格納装置を示す。ディスクブロックインデックス領域は、ファイルを構成する一連の各ディスクブロックの、ファイル格納装置内での位置を示すインデックスから構成されている。

【0023】次に、ファイル属性テーブルの第2の例を図6に示す。この例は、ひとつのファイルを分割して分散配置を行う場合の例である。ファイル属性テーブルは、図5の第1の例と同様に(1)ファイル属性領域、(2)ディスクブロックインデックス領域の二つの領域から構成される。ただし、各ディスクブロックごとにファイル属性領域が存在し、その格納場所を指定している。図6に示す例では、ファイルを構成する第1のデータブロックは第1のファイルサーバがアクセス制御を行う1番のディスク装置のインデックス100番の位置に存在することを示している。以下、第2のデータブロックは第2のファイルサーバがアクセス制御を行う1番のディスク装置のインデックス200番の位置に、第3のデータブロックは第3のファイルサーバがアクセス制御を行う1番のディスク装置のインデックス300番の位置に存在することを示している。

【0024】一方、ファイルサーバ110のプログラム構成は図7に示すとおりである。通信制御プログラム211は、LCMPネットワーク900を介する伝送のインタフェース、つまりマスタ・ファイルサーバ100と

のインタフェースとなるプロセッサ間通信装置アクセスプログラム212と、プロセッサ間通信装置アクセスプログラム206から受け取ったアクセス要求をファイルアクセス制御プログラム611が解釈できるようにプロトコル変換して渡すプロセッサ間通信プロトコル制御プログラム213から構成される。ファイルアクセス制御プログラムは、プロセッサ間通信プロトコル制御プログラム208から受け取ったアクセス要求を解釈し、目的のファイルを格納する磁気ディスク装置710にアクセスするファイル格納装置アクセスプログラム612で構成される。ファイルサーバ120、130のプログラム構成も図7と全く同様である。

【0025】次に本実施例の動作について図8を用いて説明する。

【0026】クライアント計算機10、20、30のいずれかで、アプリケーションプログラム11の実行によりファイルアクセス要求またはファイルアクセスを含む処理要求が発生すると、遠隔ファイルアクセス要求プログラム13が起動され、処理要求はローカルエリアネットワーク50を介してマスタファイルサーバ100へ伝送される。ローカルエリアネットワーク50を介した通信はクライアント計算機とマスタファイルサーバ100にそれぞれ搭載された通信制御プログラム17及び201を用いて行なわれる。ファイルアクセス要求がファイルサーバ100に送られると、ファイルサーバ100の遠隔ファイルアクセス処理プログラム301が起動される。遠隔ファイルアクセス処理プログラム301では、受信した内容を解析してクライアント計算機からのファイルアクセス要求を抽出し、ファイル管理プログラム501にファイルアクセス処理要求を送る。

【0027】ファイル管理プログラム501は図9に示すように動作する。まず、ファイルアクセス要求受付プログラム504では、遠隔ファイルアクセス処理プログラム301からのファイルアクセス要求を受け付けた後、そのファイルアクセス要求が書き込みなのか読み出しなのかを判断し、書き込みであればファイル分散配置プログラム502を起動し、読み出しであれば読み出し要求スケジューリングプログラム503を起動する。ファイル分散配置プログラム502では、書き込みファイルに対するファイル属性テーブルを作成し、次に負荷情報テーブル509を参照してアクセス要求未処理数の少ないファイルサーバをファイルを格納するファイルサーバとして決定する。また、ファイルの複製を作成して他のファイルサーバに格納する場合には再び負荷情報テーブル509を参照してアクセス要求未処理数の少ないファイルサーバをファイルの複製を格納するファイルサーバとして決定する。これらファイルと複製ファイルを格納するファイルサーバの情報をサーバ識別子としてファイル属性テーブル508に記録し、負荷情報モニタリングプログラム505を起動する。読み出し要求スケジュー

リングプログラム503では、読み出すファイルに対するファイル属性テーブル508を獲得し、そこから該当ファイルが格納されているファイルサーバを割り出す。ファイルが一つのファイルサーバだけでなく、その複製が他のファイルサーバに格納されている場合には、負荷情報テーブル509を参照してファイル本体と複製のどちらを読み出すかを決定し、負荷情報モニタリングプログラム505を起動する。負荷情報モニタリングプログラム505では、アクセス対象のファイルサーバに対するアクセス要求未処理数をインクリメントすることによって負荷情報をモニタリングし、目的のファイルのどの部分をアクセスするのかわかる情報をファイルアクセス制御プログラム601に送り、ファイルアクセス制御プログラム601を起動する。

【0028】ファイルアクセス制御プログラム601は図8に示す処理フローステップ651、652、653にしたがって動作する。まずファイル格納装置識別プログラム603はファイル管理プログラム501から渡された情報を解析しマスタファイルサーバ100がアクセス制御を受け持つファイル格納装置700へのアクセスであるのか、他のファイルサーバ110～130がアクセス制御を受け持つファイル格納装置710～730へのアクセスであるのかを判断する(ステップ651)。前者である場合には、マスタファイルサーバ100のファイル格納装置アクセスプログラム604にファイル格納装置700に関する情報を渡してファイルアクセスを指示する。ファイル格納装置アクセスプログラム604はこれを受けてファイル格納装置700へのアクセスを開始する(ステップ652)。後者である場合には通信制御プログラム201の中のプロセッサ間通信プロトコル制御プログラム208にファイル格納装置に関する情報を渡し、ファイルアクセスの実行ファイルサーバを指定して要求の転送を依頼する。プロセッサ間通信プロトコル制御プログラム208はこれらの情報をL CMPネットワーク900を介して転送できるように加工して、プロセッサ間通信装置アクセスプログラム206に渡す。プロセッサ間通信装置アクセスプログラム206は、受け取ったファイルアクセス要求をL CMPネットワーク900に送り出して目的のファイルサーバへ転送する。ここではファイルサーバ110が目的のファイルサーバであるとして説明する。ファイルアクセス要求の転送先であるファイルサーバ110では、この要求をプロセッサ間通信装置アクセスプログラム212が受け取り、プロセッサ間通信プロトコル制御プログラム213に渡す。プロセッサ間通信プロトコル制御プログラム213は、これがマスタファイルサーバ100すなわち他のファイルサーバから送られたファイルアクセス要求であることを認識すると、ファイルアクセス制御プログラム611内のファイル格納装置アクセス制御プログラム612へファイル格納装置に関する情報を渡す。ファイル格納装

置アクセスプログラム612はファイル格納装置の目的のファイルに対してアクセスを行なう。

【0029】次に実施例におけるファイルの格納のしかたを図10、図11に示す。負荷情報モニタリングプログラムによってファイルサーバの負荷をモニタリングしておき、ファイル書き込み時に最も負荷の軽い二つのファイルサーバにてファイルとその複製ファイルを格納する場合には、図10示すように同一内容のミラーファイルが格納される二つのファイル格納装置の組合せは一定ではなくなる。一方、ファイルとその複製ファイルを格納する二つのファイルサーバのペアを常に固定し、もって図11の様にファイル格納装置のペアのファイルデータを完全にミラー構成とすることもできる。この場合も、組み合わされたファイルサーバの複数のペアの間でいずれのペアの負荷が軽いかを判定して各ファイルを格納するファイルサーバのペアを決定することができる。これらに代えて、すべてのファイルサーバが同一内容のファイルデータを互いに重複して格納する様に構成することもできる。

【0030】以上、本実施例によれば、負荷情報モニタリングプログラム505によってまだ処理が終わっていないアクセス要求の個数を管理し、各ファイルサーバに対するファイルアクセス負荷をモニタリングすることにより、ファイル分散配置プログラム502と読み出し要求スケジューリングプログラム503によってファイルアクセス負荷の少ないファイルサーバへアクセスを行なうことが可能となる。しかも、ファイルとその複製ファイルを複数のファイルサーバに格納するため同一のディレクトリやファイルに対するクライアント計算機からのアクセス要求が同時に発生しても複数のファイルサーバに分散することができるようになる。すなわち、ネットワーク上に複数のファイルサーバを並接し多数のクライアント計算機間でファイルの共有を行なう際に、複数のクライアント計算機が同一のディレクトリやファイルに同時にアクセスした場合でも、特定のファイルサーバへのアクセスの集中によるボトルネックの発生とそれに伴うスループットの低下を防ぐことができ、クライアント計算機からの高スループットのアクセスを実現できることになる。

【0031】なお、上記の実施例においては各プロセッサがファイルアクセス制御を行うファイル格納装置が各々1台の場合を示したが、各ファイルサーバに複数台のファイル格納装置を接続しアクセス制御できるような構成であっても、本実施例で示した効果と同様の効果が得られることは明らかである。

【0032】さらに、本実施例で示したファイル格納装置識別プログラム、ファイル管理プログラム、ファイルアクセス制御プログラム等の各プログラムがハードウェアで構成されていても、上述の本実施例で示した効果と同様の効果が得られることは明らかである。

【0033】本発明の別の実施例を図12～図17にそれぞれ示す。図12に示す実施例はマスタファイルサーバ100のみに存在した遠隔ファイルアクセス処理プログラムを他のファイルサーバ110、120、130にもそれぞれ設けた構成である。各ファイルサーバ100、110、120、130はそれぞれLCMPネットワーク900を介して接続され、互いデータ通信が行われる。さらに各ファイルサーバ100、110、120、130はローカルエリアネットワーク50に接続される。したがってすべてのファイルサーバにおいてローカルエリアネットワーク50を介してクライアント計算機のファイルアクセス要求またはファイルアクセス要求を含む処理要求を受け付けることが可能となる。例えば、ファイルサーバ110にクライアント計算機からの処理要求が通信されると、遠隔ファイルアクセス処理プログラム311は通信内容を解釈してファイルアクセス要求を抽出し、通信制御プログラム211を起動する。通信制御プログラム211はLCMPネットワーク900を介してファイルアクセス要求をマスタファイルサーバ100に伝送する。マスタファイルサーバ100は、図1の実施例と同様にファイル管理プログラム501によりファイルを格納するファイルサーバ、もしくは読みだしを行うファイルサーバを決定する。

【0034】図13に実施例は、図1の実施例において特定のファイルサーバ100のみにファイル管理プログラムが存在する構成に代えて、ファイル管理プログラムをすべてのファイルサーバに設けた構成である。したがって、ファイルサーバ100、110、120、130の間には、マスタ、スレーブの区別はない。さらに、各ファイル格納装置700、710、720、730の記憶領域は、それぞれ4分割される。分割された領域のうち1-1、1-2、1-3、1-4は第1のファイルサーバ100に設けたがファイル管理プログラム501が管理する領域である。また、領域2-1、2-2、2-3、2-4は第2のファイルサーバ110に設けたがファイル管理プログラム511が、領域3-1、3-2、3-3、3-4は第3のファイルサーバ120に設けたがファイル管理プログラム521が領域4-1、4-2、4-3、4-4は第4のファイルサーバ130に設けたがファイル管理プログラム531がそれぞれ管理する領域である。つまり、各ファイルサーバは、互いに他のファイルサーバがアクセス制御を受け持つ行うファイル格納装置上に自身が管理可能な領域をもつ。第1のファイルサーバ100のみに設けられた遠隔ファイルアクセス処理プログラム301は、抽出したファイル書き込み要求を各ファイルサーバのファイル管理プログラムに順次振り分ける機能を有する。本実施例によれば、ファイルサーバシステム内のファイルを4台のファイルサーバで分散管理することができ、1台のファイルサーバシステム内で4組までのファイル管理を同時並列に実行す



ることが可能となる。したがって本発明の高速アクセスの効果を得ると同時に、より一層の負荷分散が可能となり、処理の並列度を上げてファイルサーバシステムのスループットを向上できるという効果も得られる。

【0035】図14に示す実施例は図13に示した実施例において、第1のファイルサーバ100に、のみに存在した遠隔ファイルアクセス処理プログラムをすべてのファイルサーバに設けた構成である。本実施例によれば、ファイルの分散管理による負荷分散と並列性の向上によりスループットを向上できるという効果が得られると同時に、すべてのファイルサーバにおいてLANを介したファイルアクセス要求またはファイルアクセス要求を含む処理要求を受け付けることが可能となるという効果が得られる。

【0036】図15に示す実施例は、各ファイルサーバ間にLCMPネットワークを設けずに、ファイルサーバ間の通信をローカルエリアネットワーク50で代用する擬似的な疎結合マルチプロセッサ構成のファイルサーバシステムである。各ファイルサーバ100、110、120、130の内部構成は図14に示した実施例のと同様である。本実施例のファイルサーバシステムにおいても、以上述べた第1の実施例の各変形実施例で得られる効果と同様の効果が得られることが明らかである。

【0037】図16に示す実施例は、擬似的な疎結合マルチプロセッサ構成のファイルサーバシステムが、ファイルアクセス要求を各ファイルサーバへ配送する機能を持つブリッジ装置60を介してローカルエリアネットワーク50へ接続される構成をとるものであり、図15のに実施例と同様の機能を実現した実施例である。各ファイルサーバ100、110、120、130の間の通信はブリッジ装置60の内部のネットワークを用いて行い、他の計算機システムとの通信の場合にブリッジ装置を介して行う。各ファイルサーバの負荷状況はブリッジ装置60でモニタリングし、クライアント計算機からのファイルアクセス要求をブリッジ装置が受信すると負荷状況をもとにファイルアクセス要求を送信するファイルサーバを選定する。本実施例のファイルサーバシステムにおいても、先に述べた図14の実施例ので得られる効果と同様の効果が得られる。

【0038】図17に示す実施例は密結合マルチプロセッサ構成のファイルサーバシステムにおいて、図14に示した実施例と同様の機能を実現した実施例である。本実施例のファイルサーバシステムはプロセッサ間通信手段としてシステムバス80を用い、これを介してファイルサーバ間の通信を行うものである。クライアント計算機との通信はシステムバス80に接続されたネットワーク通信手段により行う。本実施例のファイルサーバシステムにおいても、先に述べた各実施例で得られる効果と同様の効果が得られる。

【0039】

【発明の効果】本発明によれば、ファイルアクセス負荷の少ないファイルサーバへアクセスを行なうことができる。しかも、ファイルとその複製ファイルを複数のファイルサーバに格納するため同一のディレクトリやファイルに対するクライアント計算機からのアクセス要求も複数のファイルサーバにその負荷状況に応じて分散することができる。したがって、ネットワーク上に複数のファイルサーバを並接し多数のクライアント計算機間でファイルの共有を行なうシステムで、複数のクライアント計算機が同一のディレクトリやファイルに同時にアクセスした場合でも、特定のファイルサーバへのアクセスの集中によるボトルネックの発生とそれに伴うスループットの低下を防ぐことができ、クライアント計算機からの高スループットのアクセスを実現できる。

【図面の簡単な説明】

【図1】本発明の実施例の全体構成を示すブロック図である。

【図2】実施例の主要部の詳細構成を示すブロック図である。

【図3】実施例のマスタファイルサーバのプログラム構成を示すブロック図である。

【図4】実施例のファイル管理プログラムの構成を示すブロック図である。

【図5】実施例におけるファイル属性テーブルの一例を示す概念図である。

【図6】実施例におけるファイル属性テーブルの別の例を示す概念図である。

【図7】実施例の他のファイルサーバのプログラム構成を示すブロック図である。

【図8】実施例におけるファイルアクセス処理の流れを示すフローチャートである。

【図9】実施例におけるファイル管理プログラムの処理の流れを示すフローチャートである。

【図10】実施例におけるファイル格納形態の一例を示す概念図である。

【図11】実施例におけるファイル格納形態の別の例を示す概念図である。

【図12】本発明の別の実施例を示すブロック図である。

【図13】本発明のさらに別の実施例を示すブロック図である。

【図14】本発明のさらに別の実施例を示すブロック図である。

【図15】本発明のさらに別の実施例を示すブロック図である。

【図16】本発明のさらに別の実施例を示すブロック図である。

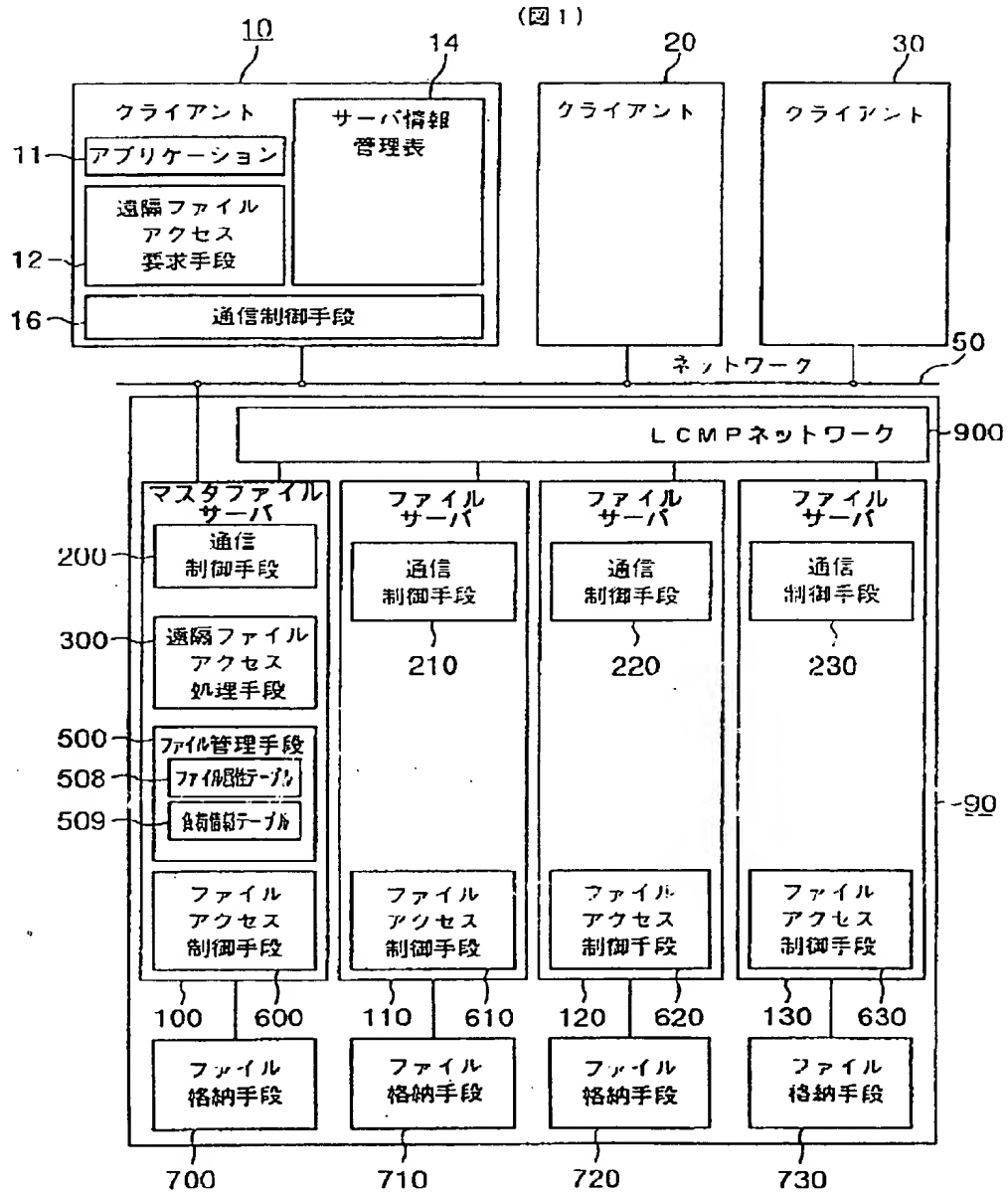
【図17】本発明のさらに別の実施例を示すブロック図である。

【符号の説明】

10、20、30…クライアント計算機、11…アプリケーションプログラム、50…ネットワーク、90…ファイルサーバシステム、100…マスタ・ファイルサーバ、110、120、130…ファイルサーバ、200、210、220、230…通信制御手段、300…

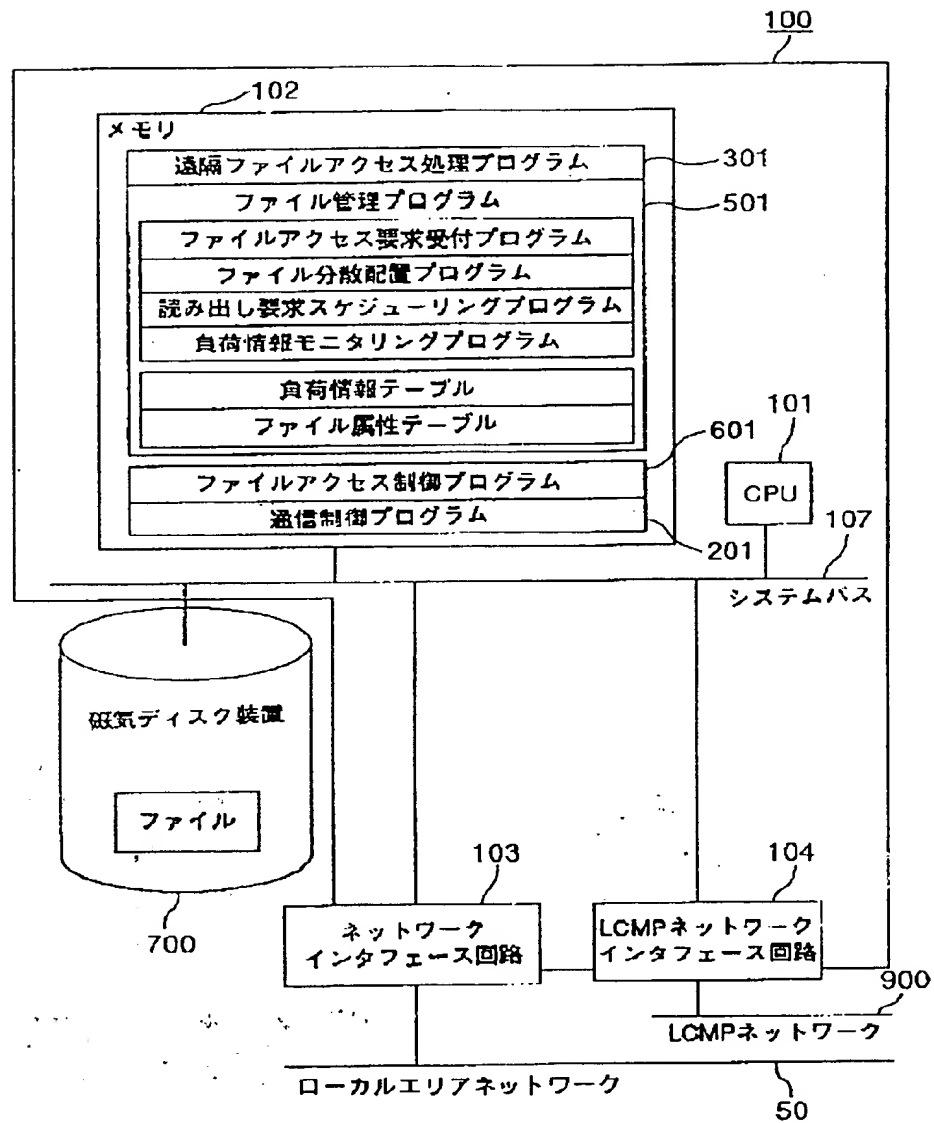
遠隔ファイルアクセス処理手段、500…ファイル管理手段、508…ファイル属性テーブル、509…負荷情報テーブル、600、610、620、630…ファイルアクセス制御手段、700、710、720、730…ファイル格納手段、900…LCMPネットワーク。

【図1】



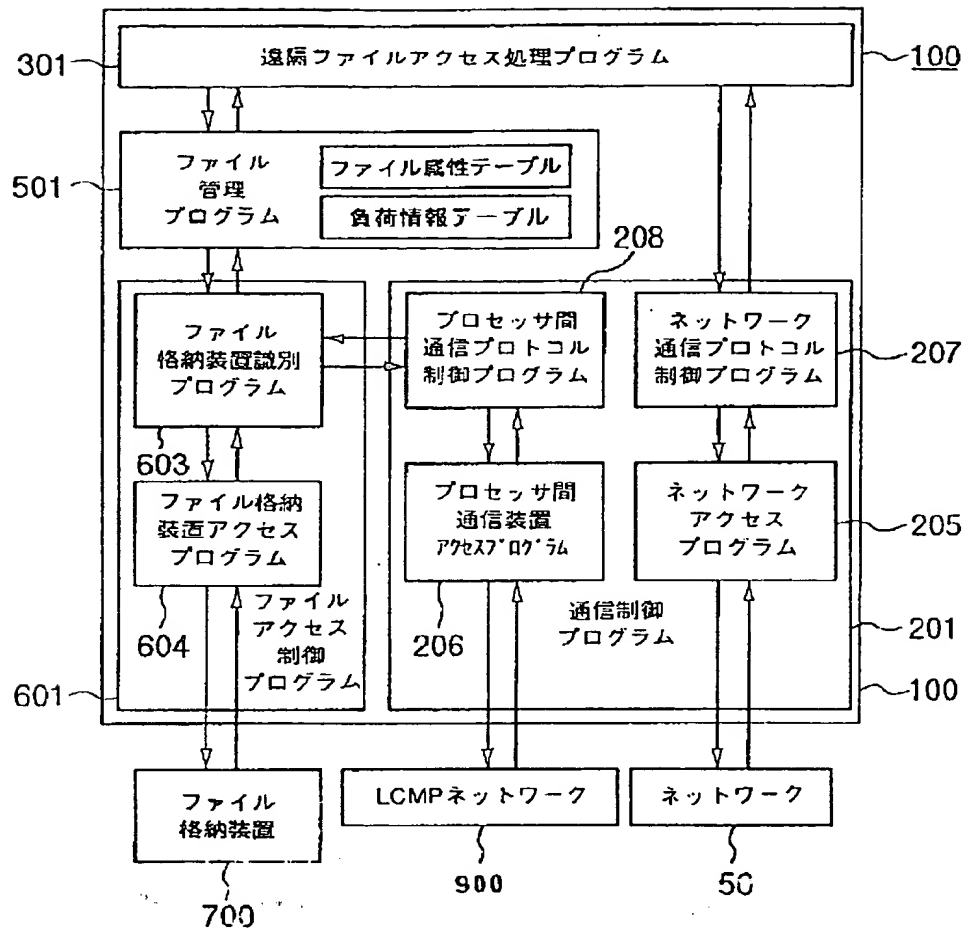
【図2】

(図2)



【図3】

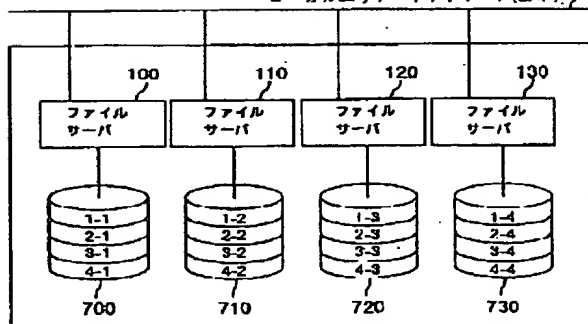
(図3)



【図15】

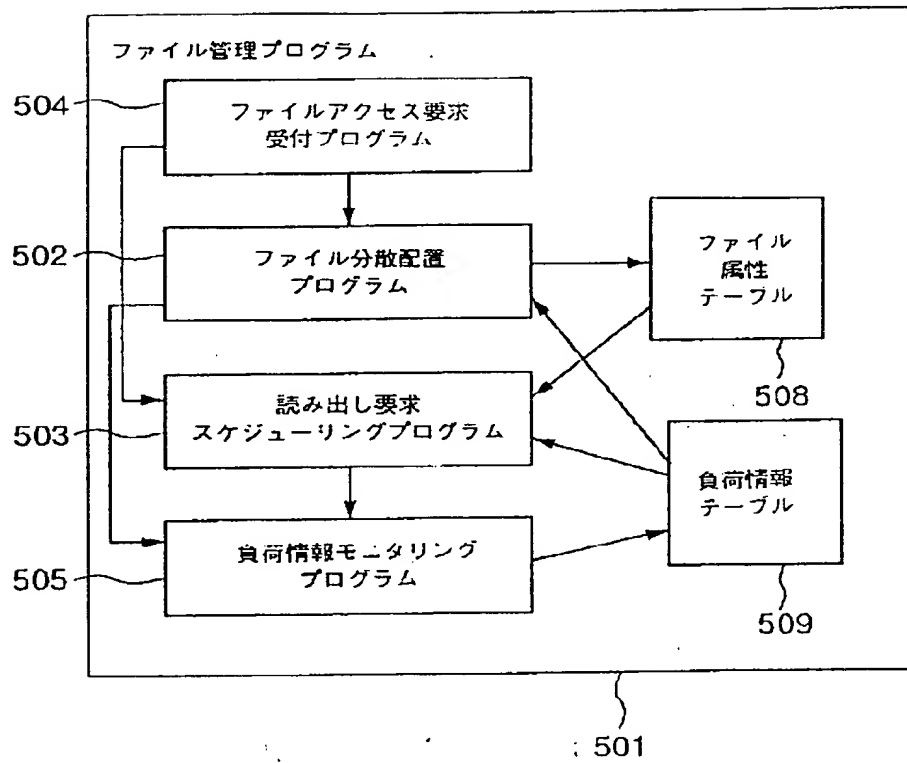
図 15

ローカルエリア・ネットワーク(LAN) 50



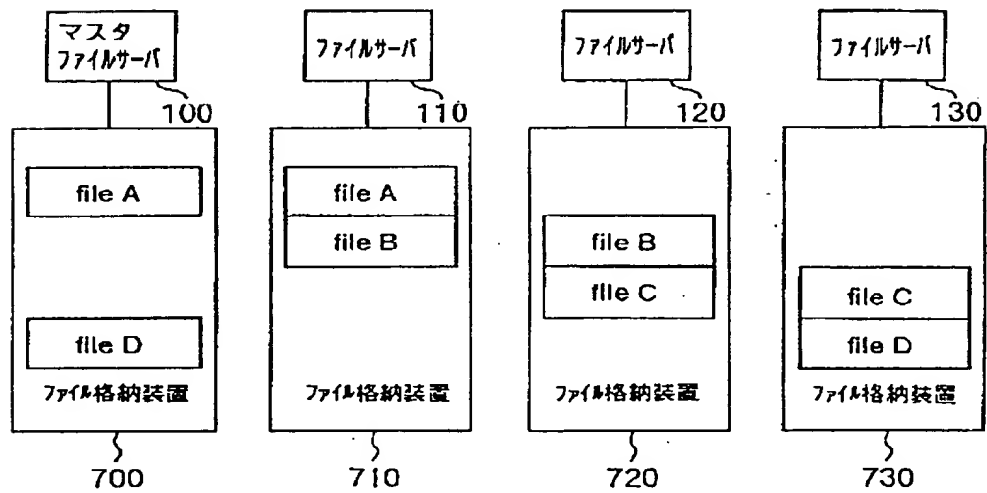
【図4】

図 4



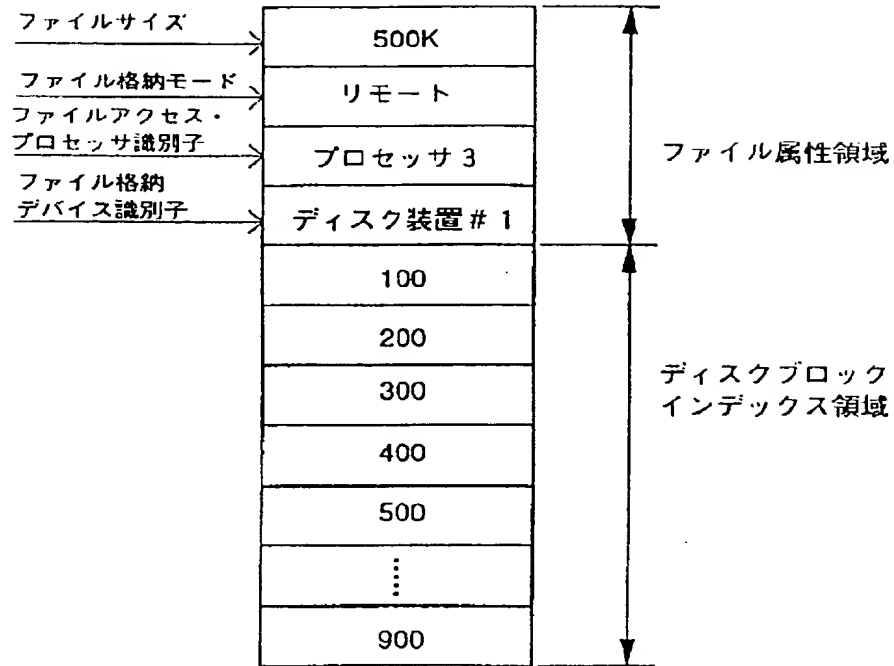
【図10】

図 10



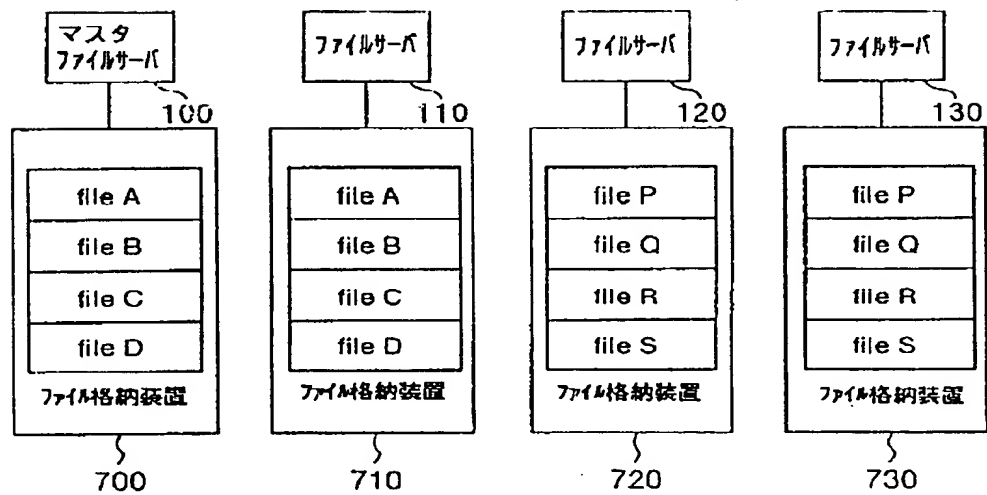
【図5】

図5



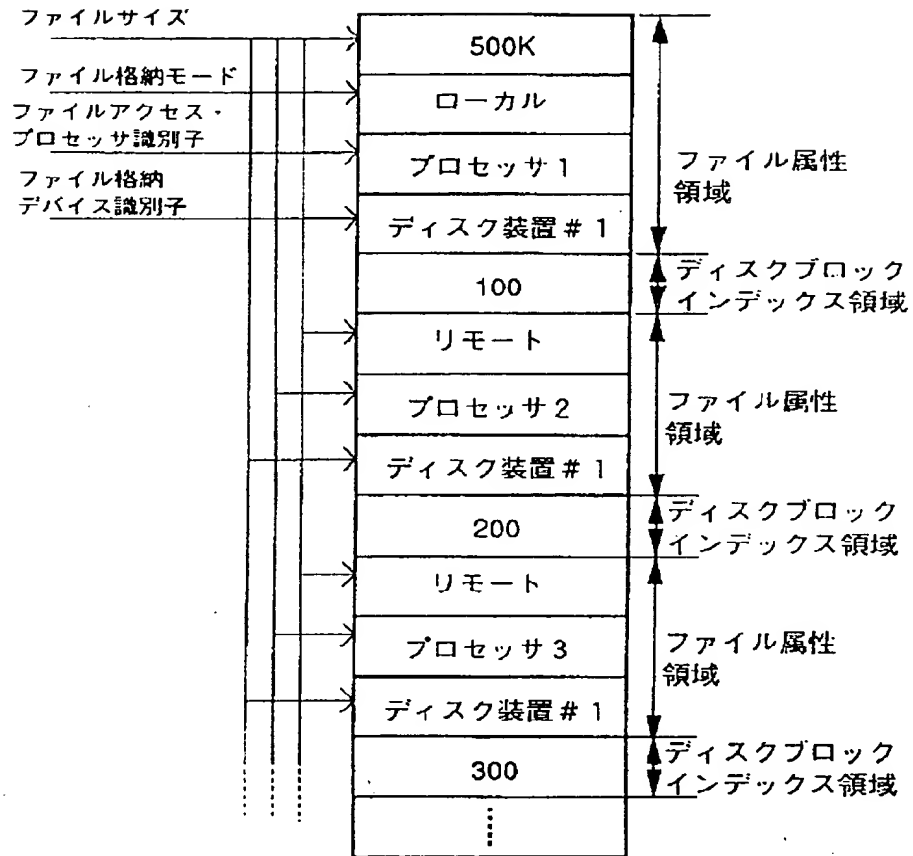
【図11】

図11



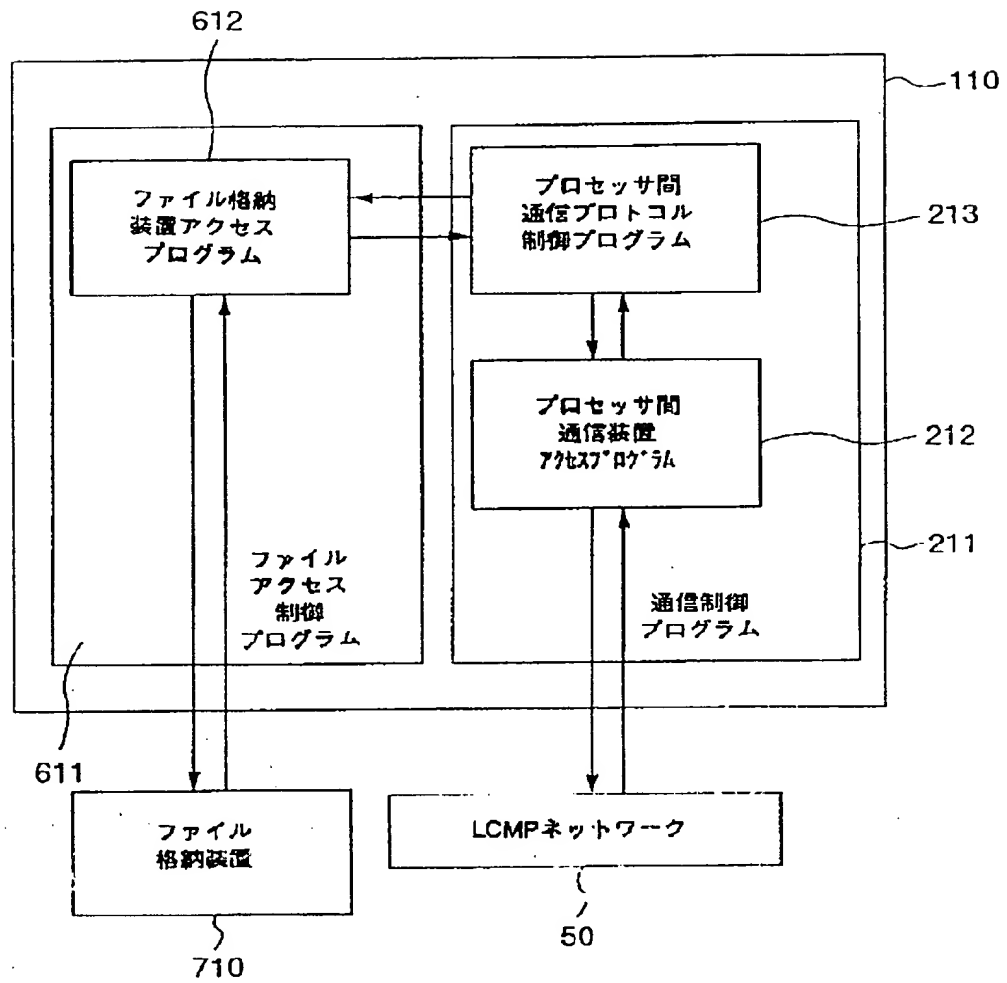
【図6】

図 6



【図7】

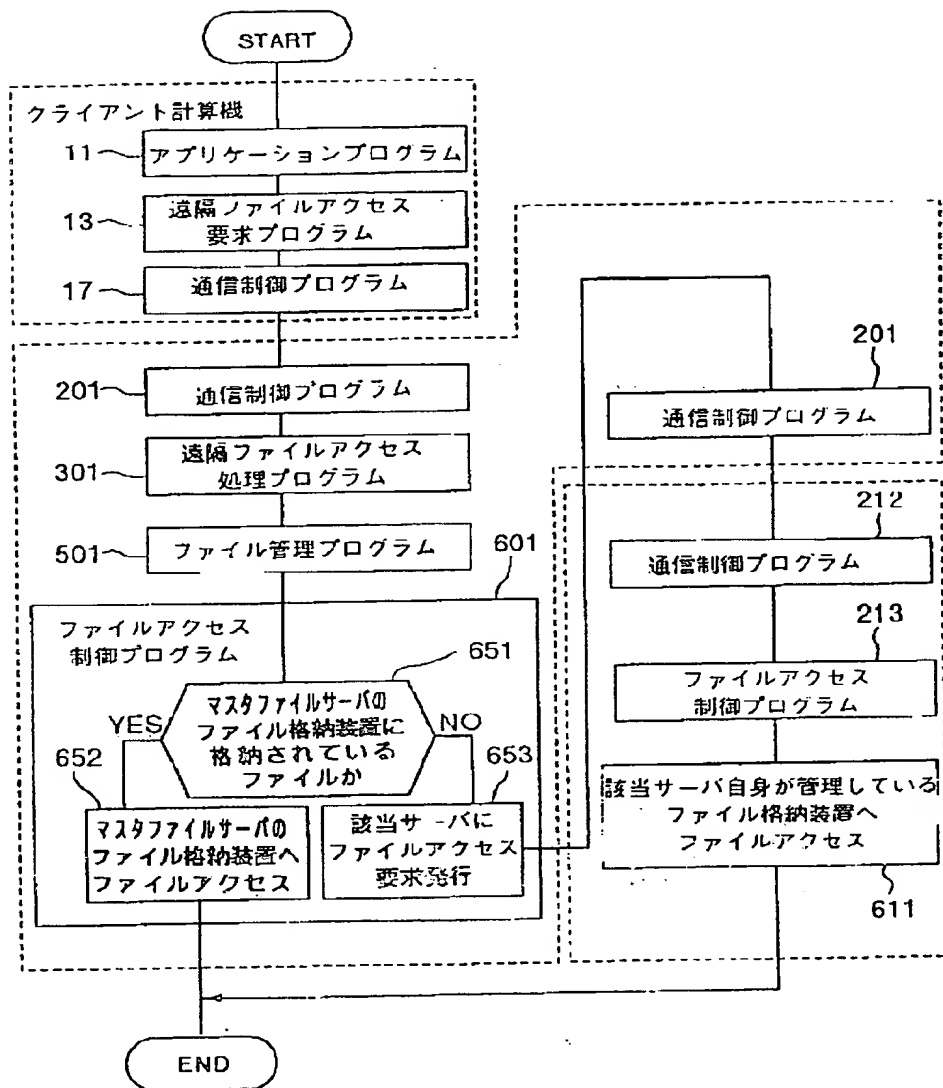
図 7





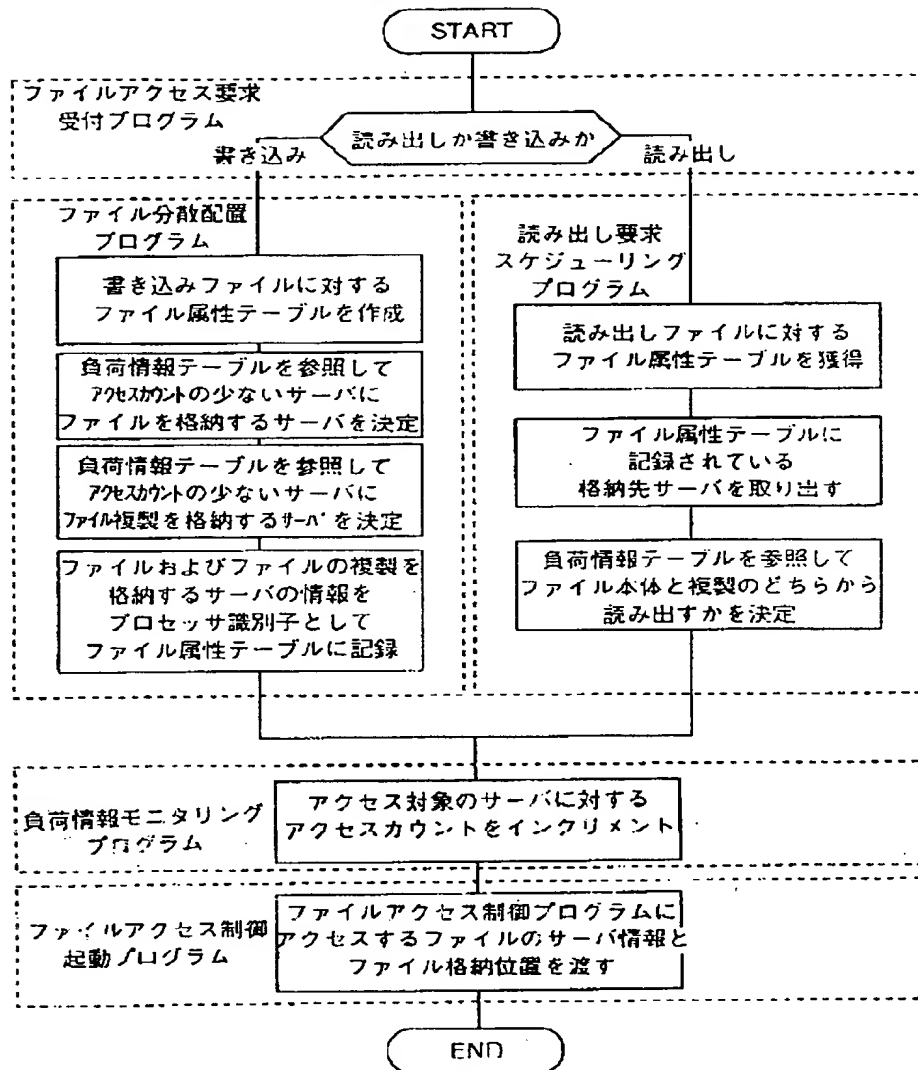
【図8】

図 8

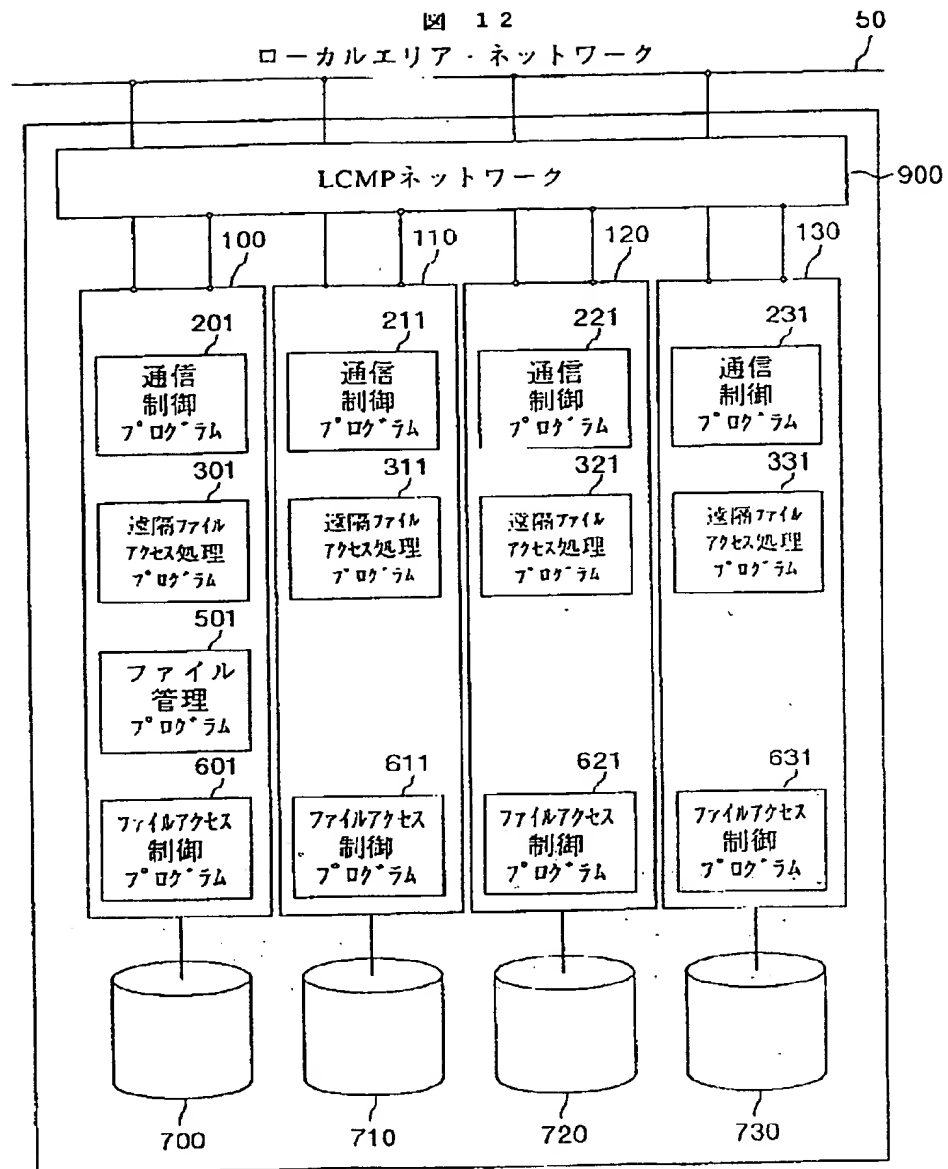


【図9】

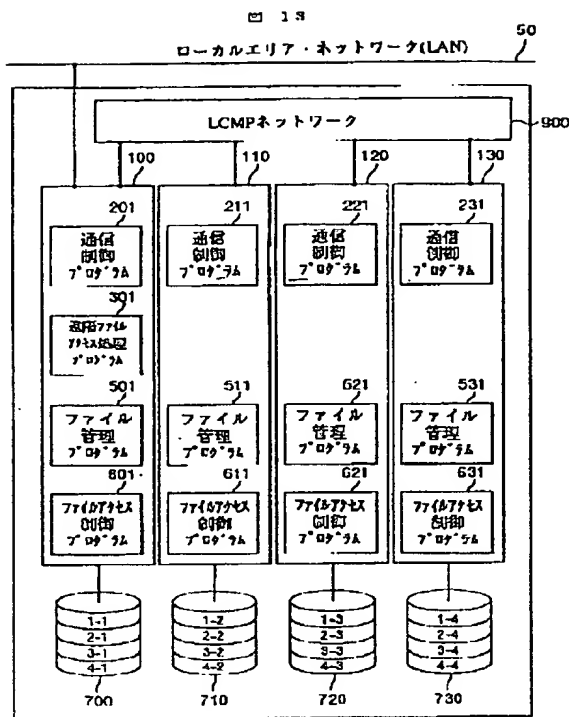
図9



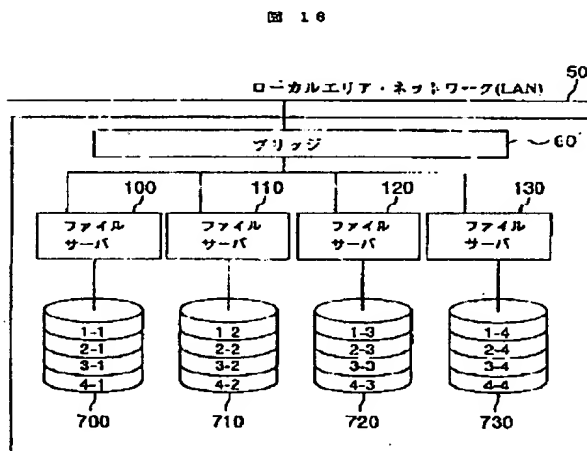
【図12】



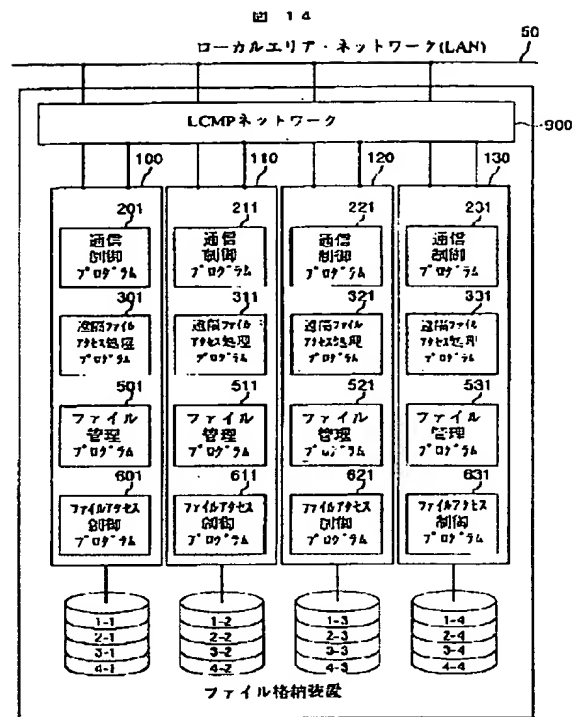
【図 13】



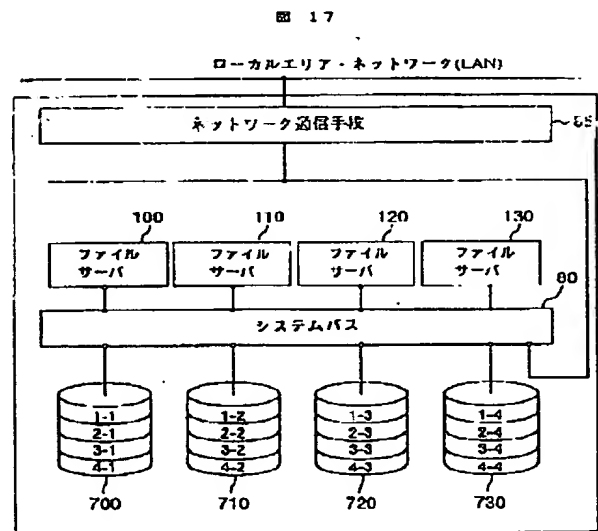
【図 16】



【図 14】



【図 17】



フロントページの続き

(72) 発明者 山下 洋史  
東京都国分寺市東恋ヶ窪 1 丁目 280 番地  
株式会社日立製作所中央研究所内  
(72) 発明者 多田 勝己  
東京都国分寺市東恋ヶ窪 1 丁目 280 番地  
株式会社日立製作所中央研究所内  
(72) 発明者 川口 久光  
東京都国分寺市東恋ヶ窪 1 丁目 280 番地  
株式会社日立製作所中央研究所内

(72) 発明者 加藤 寛次  
東京都国分寺市東恋ヶ窪 1 丁目 280 番地  
株式会社日立製作所中央研究所内  
(72) 発明者 鬼頭 昭  
神奈川県横浜市戸塚区戸塚町 5030 番地 株  
式会社日立製作所ソフトウェア開発本部内  
(72) 発明者 山田 秀則  
神奈川県秦野市堀山 1 番地 日立コンピ  
ュータエンジニアリング株式会社内

**THIS PAGE BLANK (USPTO)**